

# USO DE DESCOMPOSICIONES ATÓMICAS PARA LA MEJORA DEL MODELADO SINUSOIDAL EN CODIFICACIÓN DE AUDIO

*Pedro Vera<sup>1</sup>, Nicolás Ruiz<sup>1</sup>, Manuel Rosa<sup>2</sup>, José Curpián<sup>1</sup>, José Piñeiro<sup>2</sup>*

<sup>1</sup>Departamento de Electrónica,  
E.U.P. Linares, Universidad de Jaén.  
23700 Linares – Jaén  
E-mail: {pvera,nicolas,curpian}@ujaen.es

<sup>2</sup>Dpto. de Teoría de la Señal y Comunicaciones,  
Escuela Politécnica, Universidad de Alcalá.  
28871 Alcalá de Henares – Madrid  
E-mail: manolo.rosa@uah.es

## ABSTRACT

In this paper, matching pursuit algorithm is applied to perform atomic decompositions with sinusoidal basis on audio segments. Sinusoidal modeling usually is based on spectral peak picking by searching the FFT's maximum in each iteration. This searching go fast but has amplitude and phase errors, therefore it is necessary too many iteration to get a noisy residual. We propose use atomic decomposition based on matching pursuit to calculate sinusoidal modeling. This algorithm gets reduce the iterations to obtain similar noisy residual. However, the complexity increase. With this method we minimize number of tones in an audio segment.

## I. INTRODUCCIÓN

La codificación de señales musicales con calidad CD se ha convertido en una tecnología clave para el desarrollo de sistemas eficientes de almacenamiento y transmisión de información audiovisual. El objetivo básico que persiguen la mayoría de los codificadores de audio es conseguir codificación transparente de señales musicales en formato CD (705.6 kbps por canal) con tasas binarias lo más reducidas posibles.

La mayoría de los codificadores de audio se apoyan en dos principios básicos: 1) Eliminación de la redundancia estadística presente en la señal, y 2) Aprovechamiento del fenómeno de enmascaramiento del sistema auditivo humano. Las técnicas de codificación por transformada o en subbandas tradicionales han sido las herramientas habituales para la codificación basada en estos principios. Varias de estas técnicas han contribuido al desarrollo de los estándares de codificación de audio ISO/MPEG.

Recientemente han aparecido un conjunto de codificadores que utilizan, al menos como parte del proceso de análisis de la señal de audio, un estudio paramétrico de la señal. Así, hay codificadores [3] que extraen las frecuencias de mayor energía dentro de cada segmento de audio, realizando un tratamiento convencional del residuo. En otros casos, se sustituyen algunas bandas de la transformada por ruido sintético [2], herramienta incluida incluso en el estándar MPEG2-AAC.

El uso de descomposiciones atómicas con distintas bases puede ser una herramienta interesante para parametrizar la señal de audio [1] y conseguir, de esta manera, reducir al máximo el bit rate necesario para conseguir una calidad transparente. En este artículo se presenta el desarrollo del algoritmo matching pursuit para implementar la descomposición atómica con bases sinusoidales. De esta manera se consiguen evitar los errores

de fase y amplitud que aparecen en el estudio tonal tradicional basado en la búsqueda de máximos en la FFT de la señal.

## II. MODELADO SINUSOIDAL TRADICIONAL.

El modelado sinusoidal se basa en la extracción dentro de cada segmento de las componentes espectrales significativas de la señal. Como resultado se obtienen un conjunto de tonos denominados parciales. Cada parcial está definido por su amplitud, frecuencia y fase. La señal de entrada se puede descomponer como suma de la parte tonal y un residuo que generalmente es la parte ruidosa de la señal

$$x[n] = \sum_{k=1}^N A_k \cdot \cos(2\pi f_k n + \phi_k) + r[n] \quad (1)$$

Donde N es el número de parciales significativos considerados en cada segmento de audio.

Para extraer este conjunto de parciales tradicionalmente se utiliza la extracción de picos espectrales. Esta extracción no es más que un algoritmo iterativo. En cada iteración se busca el máximo de la amplitud FFT de la señal de entrada. Para poder distinguir picos sinusoidales próximos es necesario sobremuestrear la FFT de la señal. Además, para no confundir parciales con lóbulos laterales producidos por el enventanado de la señal, se utilizan ventanas que minimizan la aparición de lóbulos, como la ventana de Hamming. De esta forma en la primera iteración se parte de la FFT sobremuestreada de la señal enventanada, se calcula el valor del máximo del módulo, donde se extrae la frecuencia, amplitud y fase del primer parcial. Se resta, en frecuencia, un tono de estas características, siendo el resultado la entrada de las siguientes iteraciones. Así en cada iteración se obtiene un tono y un residuo de potencia cada vez menor.

La consideración de un parcial como significativo depende de la implementación que haga el codificador. Así, se pueden tener en cuenta distintos criterios. Uno de ellos es suponer significativos los tonos perceptualmente importantes [2], o simplemente, que el residuo que se obtenga sea una señal aleatoria con propiedades estadísticas ruidosas.

En la literatura aparecen algoritmos de extracción tonal algo más complejos. En [3] se propone multiplicar la señal por una serie de M ventanas ortogonales, y obtener M transformadas diferentes, extrayendo como tono el máximo común de todas ellas. Con esto se evita detectar un lóbulo lateral como parcial, por las características de las ventanas utilizadas. Sin embargo, no se soluciona uno de los problemas que aparecen: el error de fase y amplitud en cada parcial. Esto ocurre porque cada parcial se calcula como el valor de amplitud y fase de la FFT sin mayor tratamiento.

Como consecuencia, la parte no tonal de la señal, afecta a la extracción de tonos, variando el valor de amplitud y fase de la FFT con respecto al parcial óptimo. Este efecto se incrementa con el número de iteraciones.

### III. APLICACIÓN DEL ALGORITMO MATCHING PURSUIT CON BASES SINUSOIDALES.

Una descomposición atómica es la descomposición de una señal en un conjunto de bases, donde el diccionario de bases utilizado es mayor que el mínimo necesario. De esta forma se puede seguir la convergencia hacia la señal de un conjunto reducido de bases. El algoritmo matching pursuit es una forma de implementar una descomposición atómica. Es un algoritmo iterativo en el que en cada iteración se resta a la entrada la base que está mas correlada con la señal y, por tanto, la que produce un resto o residuo de menor energía. En cada iteración el residuo que se obtiene resulta de

$$r_{i+1}[n] = r_i[n] - \alpha_i \cdot g_i[n] \quad (2)$$

donde  $g_i$  es la base y  $\alpha_i$  la amplitud que minimiza la norma dos (energía) de la señal. La minimización de  $\|r_{i+1}\|^2$  es equivalente a encontrar la amplitud  $\alpha_i$  que maximiza  $|\alpha_i|^2 = |\langle g_i, r_i \rangle|^2$  para todas las bases, donde se suponen que las bases tienen norma 1. El cálculo de la correlación con todas las bases en cada iteración se debe hacer a partir del cálculo anterior para reducir la complejidad a partir de la expresión

$$\langle g, r_{i+1} \rangle = \langle g, r_i \rangle - \alpha_i \langle g, g_i \rangle \quad (3)$$

por lo que sería deseable guardar las correlaciones cruzadas entre las bases del diccionario. Tras realizar N iteraciones del algoritmo, la señal se puede descomponer en

$$x[n] = \sum_{i=1}^N \alpha_i \cdot g_i + r[n] \quad (4)$$

Para bases sinusoidales el algoritmo consistiría en encontrar el valor de  $\alpha$  máximo para cada frecuencia y fase posibles. Se puede minimizar el coste computacional considerando como bases exponenciales complejas. Así, en cada iteración se resta al residuo la base exponencial y su compleja conjugada (puesto que la entrada es real) que minimizan la energía del resto. El resto en cada iteración se calcula como

$$r_{i+1}[n] = r_i[n] - \alpha_i \cdot g_i[n] - \alpha_i^* \cdot g_i^*[n] \quad (5)$$

$$r_{i+1}[n] = r_i[n] - 2\text{Re}\{\alpha_i \cdot g_i[n]\} \quad (6)$$

Para este caso  $\alpha_i$  es un valor complejo que incluye la fase de la exponencial. Su cálculo queda, para cada frecuencia, como

$$\alpha_i = \left( \frac{1}{1 - |\langle g_i, g_i^* \rangle|^2} \right) \left( \langle g_i, r_i \rangle^* - \langle g_i, g_i^* \rangle^* \cdot \langle g_i, r_i \rangle \right) \quad (7)$$

Minimizar la norma de  $\|r_{i+1}\|^2$  es equivalente a maximizar la expresión

$$\max \left[ 4\text{Re}\{\alpha_i \langle g_i, r_i \rangle\} - 2|\alpha_i|^2 - 2\text{Re}\{\alpha_i^2 \langle g_i, g_i^* \rangle\} \right] \quad (8)$$

con respecto a  $\alpha_i$ . Se disminuye el coste computacional puesto que hay que calcular  $\alpha_i$  por cada frecuencia, y no por cada fase. Además, el cálculo de la correlación de la señal

con las bases exponenciales complejas se reduce a realizar una FFT. Tras N iteraciones la señal se descompone ahora en

$$x[n] = \sum_{i=1}^N 2\text{Re}\{\alpha_i \cdot g_i\} + r[n] \quad (9)$$

donde  $g_i[n]$  es la base exponencial (frecuencia) en la iteración i,  $\alpha_i$  es su amplitud y fase y  $r[n]$  es el residuo tras N iteraciones.

### IV. COMPARACIÓN DE RESULTADOS Y CONCLUSIONES.

Se realiza una comparativa aplicando matching pursuit y el método clásico de detección de picos con un tamaño de segmento de 2048 muestras y 8192 frecuencias distintas. Los resultados se muestran en la tabla 1.

Nº de tonos por segmento	10	20	30	40	50	100
E2/E1	5.2	6	6.5	8.2	13.4	40.7

Tabla 1. Nº de tonos por segmentos y energía del residuo del método tradicional dividida entre el método propuesto.

Conforme crece el nº de tonos extraídos en cada segmento la división entre energías del residuo aumenta. Esto ocurre por el error de amplitud y fase en el método tradicional. Sin embargo, el método propuesto determina en cada iteración el parcial que minimiza la energía del residuo. El inconveniente radica en un aumento del coste computacional en un 110%, para el mismo número de iteraciones. Pero, para llegar a un residuo de la misma potencia, el número de tonos aumenta en el método tradicional, en media, en un 155% para 50 parciales por segmento. Este valor crece con el número de parciales. Por tanto, para llegar a un residuo de similar potencia el coste computacional incluso se reduce algo (para 50 parciales por segmento).

Posteriores estudios han de incluir información psicoacústica en el proceso de extracción de tonos similares a los propuestos en [2], para corroborar la validez del método propuesto en un codificador basado en modelado sinusoidal de la señal.

### REFERENCIAS

- [1] Goodwin M.M. "Adaptative signal models: theory, algorithms and audio applications". Kluwer Academic Publishers, 1998.
- [2] Levine S., Verma T, Smith J. "Multiresolution sinusoidal modeling for wideband audio with midifications". ICASSP 1998.
- [3] Hamdy K.H, Ali M., Tewfik A.H. "Low bit rate high quality audio coding with combined harmonic and wavelet representations", ICASSP. 1996.